

R Basics

Augustin Luna
17 July, 2016

Research Fellow
Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute



What is R?

- Free, open source
- Started in 1993
- Geared towards scientific computing
 - Created by Ross Ihaka and Robert Gentleman (statisticians)
- Interpreted; similar to Python and MATLAB

First Script: Hello World!

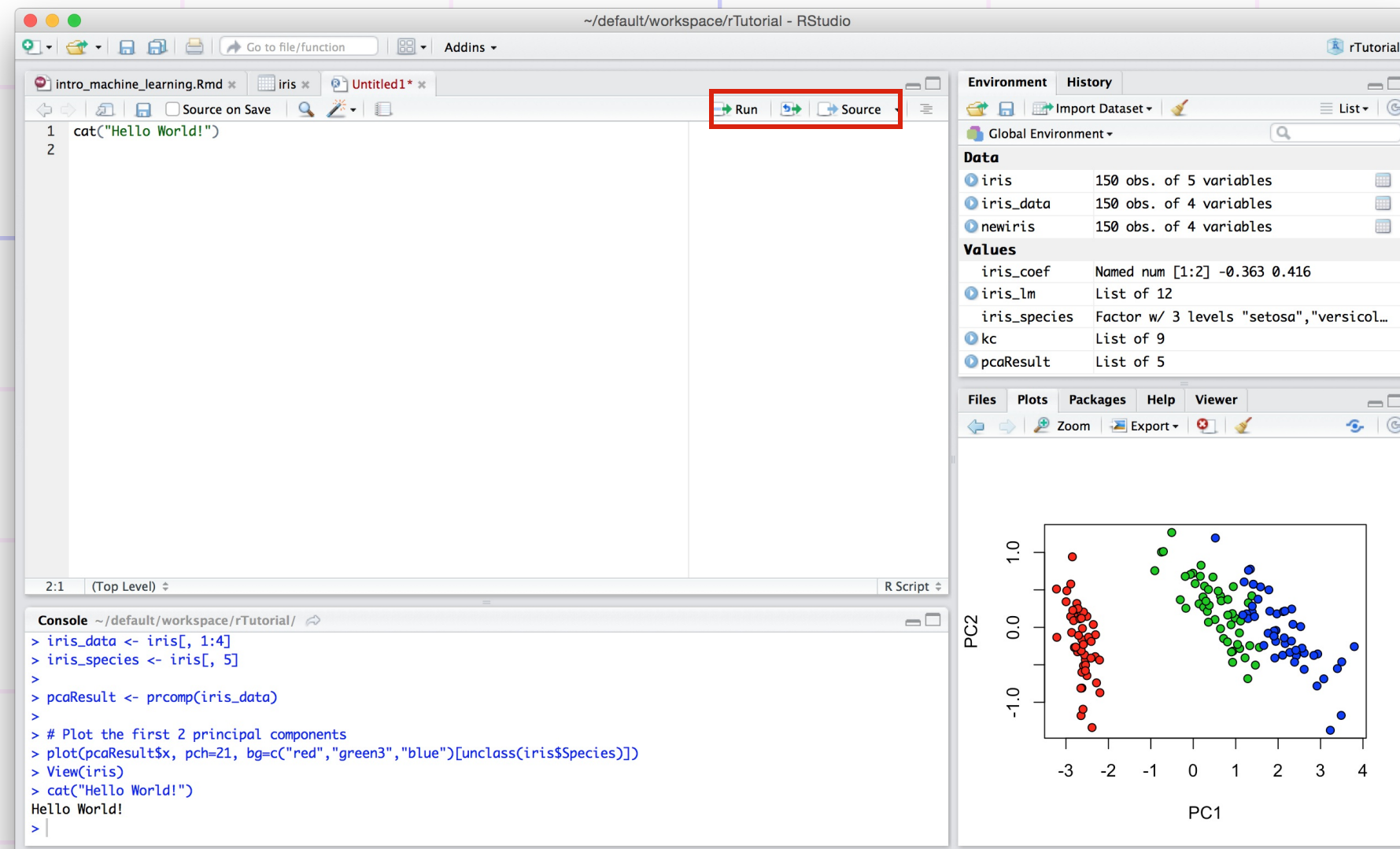
- `cat()` prints a simple message in the console

```
cat("Hello World!")
```

```
Hello World!
```

Running Hello World Script

- “Run” button runs current line or selected lines
- “Source” button runs all lines in file



RStudio Overview

~/default/workspace/rTutorial - RStudio

intro_machine_learning.Rmd x

Knit HTML Addins Run

```
51
52 Compare the Species label with the clustering result
53
54 ```{r}
55 table(iris$Species, kc$cluster)
56 ```
57
58 Plot the clusters and their centres. Note that there are four dimensions in the data and that only the first two
59 dimensions are used to draw the plot below. Some black points close to the green centre (asterisk) are actually
60 closer to the black centre in t
61
62 ```{r}
63 plot(iris[c("Sepal.Length", "Sepal.Width"), kc$cluster], pch=21)
64 points(kc$centers[,c("Sepal.Length", "Sepal.Width"), "blue"), pch=8, cex=2)
65
66 ## Dimension Reduction
67 ### Principal Component Analysis
68
69 From: http://www.r-bloggers.com/computing-and-visualizing-pca-in-r/
70
71 ```{r}
72 iris_data <- iris[, 1:4]
73 iris_species <- iris[, 5]
74
75
76 Machine Learning Introduction R Markdown
```

Editor

Environment **History**

Global Environment

Data

iris	150 obs. of 5 variables
iris_data	150 obs. of 4 variables
newiris	150 obs. of 4 variables

Values

iris_coef	Named num [1:2] -0.363 0.416
iris_lm	List of 12
iris_species	Factor w/ 3 levels "setosa","versicol...
kc	List of 9
pcaResult	List of 5

Files **Plots** **Packages** **Help** **Viewer**

Zoom Export

PC2

PC1

Console ~/default/workspace/rTutorial/

```
> points(kc$centers[,c("Sepal.Length", "Sepal.Width"), col=c("red", "green3", "blue"), pch=8, cex=2)
>
> # Chunk 7
> iris_data <- iris[, 1:4]
> iris_species <- iris[, 5]
>
> pcaResult <- prcomp(iris_data)
>
> # Plot the first 2 principal components
> plot(pcaResult$x, pch=21, bg=c("red", "green3", "blue")[unclass(iris$Species)])
>
```

Top 20 Packages (kdnuggets.com)

Package	Downloads
Rcpp	693288
ggplot2	598484
stringr	543434
plyr	523220
digest	521344
reshape2	483065
colorspace	476304
RColorBrewer	453858
manipulate	395232
scales	394389

Package	Downloads
labeling	373374
proto	369096
munsell	368949
gtable	364015
dichromat	362562
mime	352780
RCurl	340530
bitops	322743
zoo	302052
knitr	295528